My title is "what is a gene ?" but what I really want to talk about is some recent work on the fine-structure of the gene; work which may lead us to understand more about the molecular basis of genetics and perhaps something about the way that genes act.

Let me first sketch for you the classical picture of the gene. You can find this in any elementary textbook of genetics. It starts by describing some simple breeding experiments - usually the historic experiments of Mendel on peas - and shows how they can be explained by a straightforward theoretical scheme. Then the complications due to "recombination" - as it is called - are introduced, and finally the reader learns how all these ideas can be understood in terms of the chromosomes, the small thread-like bodies within the nuclei of cells. The chromosomes carry the hereditary determinants, and the genes, the units of heredity, are strung out along the chromosome, in a linear order, like beads on a string.

The gene, then, is the unit of heredity. It is the building brick of the material passed on to a child from its parents, half from each of them, and it is these genes which mainly determine the sort of child it is. It is not known how many different kinds of genes there are, but each human being probably has more than 10,000 ~~of them.~~ different ones.

That is the standard picture - but I want to explain to you that the gene - the gene in the very old-fashioned sense of the indivisible unit of heredity - has been split, and split decisively. To do this I shall discuss a simpler system so that you can grasp the essential points more easily .

The system I shall describe is that of a virus that attacks bacteria. You may feel that this is rather remote from human genetics, but it is one of the astonishing discoveries of modern biology how similar the basic biochemical and genetic processes are in all living material. This virus is quite a small object by normal standards - it is too small to be seen even with a very powerful light microscope - but as viruses go it is quite large. Its molecular weight - if we may use such a term - is about two hundred million (the molecular weight of water is 18) and it is twenty or thirty times bigger than a small virus such as the poliomyelitus virus. It is made up of only two types of chemical - protein and deoxyribonucleic acid (DNA for short). Most if not all of the protein component corresponds to the body of the virus, whereas the genetic part appears to consist mainly if not entirely of DNA.

Let me say first, without giving the evidence, that the genetic material appears to be all in one long piece - in a higher organism we would say that there was one chromosome - and that there is only one copy of it in the virus. To use the genetic jargon it is haploid, whereas human beings, for example, are diploid; that is, we have <u>two</u> copies of each chromosome, one from each parent. It is the fact that the virus is haploid which simplifies the explanation. The exact number of genes in this virus is not known, but a conservative estimate would be twenty, arranged in a linear order along the chromosome.

When the virus comes into contact with a suitable bacterial cell it attaches itself by its tail (it has a polygonal head and a long thin rigid cylindrical tail). All the DNA then enters

the bacterial cell and most of the protein stays outside. Twenty minutes later the cell bursts open and a hundred or so new complete virus particles are released. Thus in this short time the system has made many copies of the infecting virus. Moreover the copies are exact copies. Various strains of the virus exist and the viruses that come out of the cell are exactly like the particular virus particle which went in. In other words the virus breeds true.

Very occasionally, however, the replicating mechanism makes a mistake, and a slightly different virus emerges. If we now infect a second time with this changed virus its descendents are all like itself; that is, of the altered form. In a higher organism such a change would be called a mutation, and there seems no reason why we should not use the word here. The virus, then, can experience a mutation, and the ~~mutation~~ mutant will breed true.

Now what will happen if one infects with two different mutants of the virus simultaneously ? It can be shown that both penetrate into the cell, both multiply inside the cell, and copies of both emerge when the cell is burst open. But the interesting fact is that, in addition, new virus strains appear whose genes come partly from one parent virus and partly from the other. The picture evoked to explain this is that during the copying process inside the cell a new copy may occasionally start to be formed on one parent and then, at some random point, skips over and finishes the process of copying on the other parent. This process is called "recombination" - again because of its similarity to a similar process in higher organisms.

It is this recombination which allows geneticists to map the genes on the chromosome. If two genes are very close together the chance of the skipping-over process taking place between them is rather small. If, on the other hand, they are far apart it will happen more often. By finding how frequently recombination occurs for any two genes one can get some idea of their distance apart, and it is by this method that geneticists have shown that genes are arranged in a linear array along the length of the chromosome.

Let us forget recombination for a moment, and think about what happens if we have a virus one of whose genes is defective. If this virus is used by itself it may be unable to reproduce. However if another strain of virus, with that particular gene intact, is used simultaneously, it may be able to do that part of the work for two, and both viruses may be able to reproduce. In fact if the second virus has a defect in a different gene, of which the first has a good copy, they may be able to help each other. A pretty picture !

Fortunately it has some technical use. For suppose both these two viruses, instead of having their defects in different genes, had a mistake in the same gene. Then neither of them would have a good copy, and reproduction might not be possible. Thus this test (of seeing whether two defective viruses can assist each other or not) shows whether the defects are in the same gene or not. This is one of the definitions of the gene, in fact - the unit of physiological action, and it is the meaning I shall be using when I use the word gene.

It is not the only definition, however. We could define the gene as the "unit of mutation". That is, the smallest part of a chromosome which can change, or mutate, independent of the other parts. Or again, we could define the gene - and indeed many geneticists do - as the unit of recombination - the smallest part that can be moved about by the recombination process: the smallest step in the skipping-over process if you like.

So there we have three defintions of the gene. The unit of action, the unit of mutation and the unit of recombination, and for many years it was felt that, in spite of some exceptions, these three definitions refered to the same thing - the gene, as it was called.

It is now becoming clear that this picture was ~~x~~ oversimplified. Let me give you the new picture straight away, to help you follow the argument. In the new picture the unit of action is a large thing, roughly ~~xxxxxxx~~ corresponding to what used to be thought of as the gene, but the other two units now appear to be very much smaller. We can put this another way. Let us for the ~~moment~~ rest of this talk use the word gene for the unit of action. Then we can summarise the new discoveries by saying that not only can geneticists map the order of genes on a chromosome, but they can map the order of the different changes within a gene. In other words, the gene - the unit of action - has been split by the genetic mapping technique.

Now I would not want you to think that this is a totally new idea, because evidence along these lines has been available for years. What has transformed the situation is the weight of

the new evidence. This has mainly come from two sources. Drs. Pontecorvo and Roper, working in Glasgow on the mould Aspergillus, and Dr. Benzer working at Purdue University, in the States, on the virus system I have described. Br. Benzer's system has a technical advantage over that of the Glasgow workers because it can be pushed more quickly to finer limits, but the work on the mould makes us think that the new data on the virus is not a freak, but likely to be of general occurence. Moreover cases are known in flies, in corn and in other organisms.

Let me briefly summarise what Dr. Benzer finds. For technical reasons he has studied one particular gene, known as r II . His first finding is that r II is not one gene, but two, next to one another, and with closely related functions. He has picked up mutants of this pair of genes on about a thousand different occasions. Sometimes he finds that a newly-arising mutant appears by his tests to be identical with one he has found before, but often it is quite different from any previous one. In all he has so far found about 250 <u>different</u> mutants in his two genes combined. In other words each of his genes has been marked (or split, if you like) in over 100 different places.

His next finding is that he can arrange all these different mutants in a <u>linear</u> order, as if they were a lot of different marks on a piece of string. There are no real exceptions to this. There are a few apparent exceptions, but these appear to be due to "deletions" as they are called - meaning the actual loss of a small piece of the string. In every such case he never finds

a reverse mutation - the lost piece of the string never reappears again, as one might expect.

As far as he can tell, allowing for the limitations of his technique, there is no space between his two genes. As soon as one ends, according to his mapping, the other begins, without any appreciable gap.

Another interesting finding is that there are mutants which he can tell are different, but which he cannot seperate on the map; they appear to occur at the same point. You might wonder how he knows they are different. The character which distinguishes them is their reverse mutation rate - the rate at which they mutate back to the original version. For example one particular mutation site is very common. He has picked it up about 130 times out of his set of 1,000 . About half of ~~these~~ this 130 have one back-mutation rate, whereas the remainder(with one exception) have another one. In fact he appears to be finding that the unit of mutation is not quite the same as the unit of recombination though both appear much smaller than the unit of physiological action.

It is not unreasonable to ask how small his smallest unit is. How near is he to the molecular limit - the molecular "graininess" if you like - of the genetic material ? It is possible to make a very rough estimate of this if certain simplifying assumptions are made. The gentic material consists mainly of DNA. Let us assume that it is wholely DNA. Now DNA, as you know, is a polymer, made up of small molecules called nucleotides joined end to end. The total number of nucleotides

in the whole virus is about 400,000 though there are reasons for beleiving that only part of this corresponds to the genetic map. Knowing roughly the total length of the genetic map, ~~and the size of the smallest interval in his fine-scale mapping~~ Benzer has calculated the number of nucleotides which correspond to his smallest map-distance. The answer, which is only approximate, comes to around half a dozen - and it might be fewer. In other words by genetic means he is dividing the gene down to, or almost down to, the molecular level. You can understand why those of us who are interested in the molecular basis of genetics find his work tremendously interesting.

Finally we may ask - what next ? It is a very reasonable speculation that each gene controls, directly or indirectly, the production of a particular protein molecule. More precisely, that it controls the order of the amino acids in one polypeptide chain of a protein molecule which the cell is producing under its influence. What we suspect is that the linear order inside a gene corresponds to the linear order of amino acids along the polypeptide chain of the relevent protein. ~~which the cell is producing~~. If we could find the protein controlled by ~~corresponding~~ to Dr. Benzer's gene we might be able to discover, with modern techniques, the change in the order of the amino acids produced by any particular mutant. If we could do this we could soon see if the two orders - the linear mapping of the gene and the linear arrangement of the amino acids - were related. Dr. Benzer is coming to work with us at Cambridge next year, and this is exactly what we shall try to do.